INTEGRATING A DIGITAL REPOSITORY SYSTEM

Martin Borchert & Joanna Richardson
Flexible Learning & Access Services
Griffith University, AUSTRALIA
m.borchert@griffith.edu.au

Abstract

As part of its commitment to excellence in student-centred learning, Griffith University has initiated the Learning@Griffith project to assist in the delivery of the University's teaching, learning and research goals. A core component of this project is to integrate a digital repository system. This paper outlines the partnership recently formed between HarvestRoad and Griffith University to implement Hive, a high-use, centralised, federated digital repository service through the provision of a range of client-focused bureaus (digital collections), search and access facilities, and workflow administration systems. Copyright management and compliance are also discussed. A number of start-up bureaus are being established including an art image collection, digitized exams, and an ePrint service. The next phase includes the transfer of teaching and learning objects for all courses. Processes will be differentiated based on whether copyright ownership is internal or external to Griffith. This paper reports on progress, future applications, and lessons learned.

Keywords

Digital repository, copyright, eprints

Introduction

The convergence between teaching and learning, library, and information communication technology professions and the provision of services is well known. University staff are learning new roles and developing new structures and systems in order to deliver new interdisciplinary cross-departmental services to clients. Information Services at Griffith University undertook a restructure in 2002 for such purposes. The department of Flexible Learning and Access Services was created to bring together content creation, provision and management supporting teaching and learning together with purchased content for print and digital library collections. The objective was to foster the provision of more highly and appropriately used content, delivered to university staff and students within a more seamless and service oriented environment.

New content delivery systems better able to manage and deliver teaching, learning and library content are required to achieve such a goal. The historical divisions between Learning Management Systems (LMS), Content Management Systems (CMS), Institutional Repository Systems (ePrints), Federated Search and URL Resolvers, Copyright Management Systems and Integrated Library Management Systems (ILMS) have impeded progress towards achieving organisational change and the realisation of new generation services.

The proliferation of digital learning objects supporting online learning and digital library content has required the development of content management systems and workflows. Griffith University uses the Blackboard LMS system behind its Learning@Griffith service. The benefits of using an LMS in course content delivery are obvious but so are the limitations in sharing and managing digital learning objects. Local digital library collections have been hosted on Web servers, which again provide very limited levels of content management functionality. Many institutions are developing Institutional Repository Systems (ePrint services) using different specialised software such as eprints.org.

Two approaches could be adopted towards resolving the disaggregated systems issue, the first being to employ new systems that cross traditional functional boundaries, and the second being the further development of standards and functionality promoting improved communication and integration between existing systems (McLean & Lynch, 2004; IMS 2003), thus acknowledging systems' current customisation towards specific functions. Long (2004) suggests that the real challenge is to identify those common elements of information systems which share services so as to facilitate the exchange of content.

The concept of digital repositories has arisen as a method for storing digital resources—with varying levels of control as outlined by Peters (2002). Leslie (2004) questions the difference between institutional repository software such as DSpace and Fedora and 'learning object repository' specialised software such as CAREO and notes that DSpace's FAQ page promotes the software as multipurpose and cross-functional.

In 2003 Griffith University selected the HarvestRoad Hive digital repository system. Griffith's aim is to use Hive for as many digital repository system applications as the flexibility of the system will allow. The current and future application of Hive in the management and provision of digital learning objects, local library collections, image collections, and institutional ePrints is discussed. The Griffith Hive project is still in its early developmental stage with no services currently offered, however the concept of applying one technical solution to varied online learning and library information management needs is explored in this paper.

Learning@Griffith

Learning@Griffith is a centralised, university-wide online teaching and learning management and delivery system built on the Blackboard LMS, within the portfolio of the department of Flexible Learning and Access Services. Learning@Griffith is currently the most heavily used single application of Blackboard 6.0.11 in the world. Through concerted implementation across all faculties and other elements of the university, Griffith has increased the number of courses (subjects) loaded to Blackboard from 1,400 in 2002 to 5,500 in 2003 and to 9,000 in 2004. In semester 2 2004 Griffith has some 5,000 lecturers assigned to courses. Usage is very high and increasing, with some 2,500 simultaneous student and staff connections recorded at peak periods during semester 1 2004. Between 2003 and 2004 the volume of data stored in Blackboard increased from 20 to 70 Gb. Semester 1 2004 content included 134,000 file objects, 180,000 content objects, and 25,000 announcements.

Academics are providing much course content via Web pages within the Blackboard environment and are loading related digital learning objects directly to the LMS. Academics currently have access to their own content only within Blackboard and are required to manage their own digital learning objects, the result being that they do not currently have an environment in which they can share and re-use learning objects. Many academics are also choosing to delete objects from previous semesters rather than hide and store objects for re-use. Together these issues are creating an environment of increasing re-creation and duplication of similar objects.

Digitised course readings of copyright materials such as book chapters and journal articles have been separated from Griffith owned copyright materials since 1999. Course readings are digitised by the Digitisation and Distribution team, a centralized service within FLAS, at the request of lecturers and in response to the inclusion of copyright readings in print dossiers or coursepacks. Digitised readings are available via links within the Learning@Griffith course pages and from the GriffLink library catalogue; the digitised readings themselves being stored on a secure web server. MARC records within GriffLink store information required for copyright management although copyright management functionality is quite limited. 9,000 digitised readings are currently available.

HarvestRoad Hive Project

Hive is being developed as the solution for as many digital repository purposes as the system will accommodate. Key components of the selection of the Hive system included:

- The opportunity to partner with the Australian vendor HarvestRoad to co-develop Hive and thus produce a product attuned to Griffith's needs
- Hive's integration and cross-functionality with the Blackboard learning management system used by Griffith
- The extensible, federated and independent nature of the Hive product
- Its ability to support multiple procedures and workflows
- Copyright and other reporting capabilities, and
- Its ability to be fully customisable in bureau structure, acceptance of any and multiple metadata schema, and thus its application to a wide variety of online learning, library and digital collection applications.

Hive is designed to store digital resources such as learning objects, images, readings and other materials. Objects can be grouped into collections or bureaus. Each bureau can be searched individually or by federated searching. Once fully implemented, the Hive will initially be used to house digital collections on behalf of the Division of Information Services, and will later be extended to store an increasing number of digital collections for faculties and other administrative elements across the University. Griffith is also a fully licensed training partner in the Australian market.

A new Digital Repository team was established in 2004 within the department of Flexible Learning and Access Services (Division of Information Services). The team currently consists of four staff including Digital Repository Administrator, Metadata Indexer and two contracted Digital Repository Project Officers. Team members were primarily selected based on a general understanding of flexible teaching and learning objectives, IT competency, proficiency in metadata schema, ability to work in a changing environment, and high-level customer service skills. To establish the required working relationships and synergies this team was located proximally to the Learning@Griffith Team at the Nathan campus.

Digital Repository Roadmap

In 2004 Information Services (Griffith) published a 2004-2006 "Roadmap" for each product and service offered by the division for the University. These roadmaps are planning documents used to frame the vision statement, services, technological, organisational and client side impacts, and to foresee financial and staffing requirements. A roadmap was produced for the digital repository and is being used as a broad project plan.

The vision statement for the digital repository is:

> Griffith Digital Repository 2004-2006: A high use, centralised, federated digital repository service supporting the teaching, learning (via Learning@Griffith) and research goals of the University through the provision of a range of client focused bureaus (digital collections), search and access facilities, and workflow administration systems providing storage and management solutions for digital objects.

Base development elements include:

- Successful implementation of the Hive software running to high performance standards
- Selection and establishment of a high performance Digital Repository team
- Development of the underlying structure of digital collections (bureaus, categories, subcategories)
- Development of general policy on granularity (and level of re-usability) of objects to be managed
- Development and implementation of a base generic minimum metadata schema to be applied to all bureaus (to facilitate federated searching)

- Development and implementation of a base generic workflow for the management of objects
- Implementation of a copyright management system

Starter Applications

- Selection and implementation of appropriate initial 'starter' projects to help develop deeper understanding of Hive, integration with Blackboard, workflow management, and to develop communication and training strategies for academic staff
- Digital Asset Management System
- Medicine collection required for the Flinders course
- QCA Digital Art Collection
- A single School's course readings
- University Exam paper collection
- Development of appropriate extension to the base metadata schema for each bureau to be established
- Development and implementation of appropriate workflows for each bureau to be established

Core Applications

The following have been identified:

Griffith owned copyright teaching and learning objects supporting Learning@Griffith

- Development of Griffith policy and guidelines on intellectual property, and ownership and reusability of objects
- Development of Griffith policy on the management and storage of non-copyright teaching & learning objects supporting Learning@Griffith
- Development of a communication and training strategy supporting the management of non-copyright teaching & learning objects supporting Learning@Griffith
- Development of bureau(s) and workflows for the management of non-copyright teaching & learning objects supporting Learning@Griffith
- Integration with Learning@Griffith and interfacing for student users, and academic administrators for teaching, learning and research purposes
- Decision to retrospectively move objects to the digital repository or not (pending funding)

Copyright teaching and learning objects supporting Learning@Griffith

- Development of Griffith policy on the management and storage of copyright teaching & learning objects supporting Learning@Griffith
- Development of a communication and training strategy supporting the management of copyright teaching & learning objects supporting Learning@Griffith
- Development of bureau(s) and workflows for the management of copyright teaching & learning objects supporting Learning@Griffith.
- Project to move all copyright materials to the appropriate bureau within the digital repository (funding permitting)
- Further development of the copyright management module within Hive such that it supports CAL reporting requirements

Institutional ePrint repository including:

- Development of Griffith policy on the management and storage of Griffith's research pre-print and post-print publications
- Development of a communication and training strategy supporting an eprint service
- Development of bureau(s) and workflows for the management of an eprint service

Extension Applications

Potential future uses are extensive but as yet unclear. The demand for additional bureau applications will become apparent over time. These may include for example:

- Additional image collections
- Streaming video collections

- Sound files
- Administrative digital assets from across the university – forms, policies otherwise stored on web servers
- Database/ejournal licenses and contracts
- Specific faculty/school/centre collections of digital resources
- Theses collections not otherwise collected, such as Honours theses, or fourth year student assignments
- Locally hosted digital library contents, ebooks, reports etc
- Research data collections (as distinct from publications)
- Commercial applications (fee based hosting)
- Other undetermined applications

Major factors affecting the development of the digital repository include future development of Hive, technological changes to Blackboard and other interfacing systems, commercial opportunities and cooperative agreements, developments in metadata schemas, the purchase of content coursepacks, high level and academic support for policies and workflow practices including academic self-deposit and metadata creation, and Information Services' ability to successfully market individual projects such as ePrints.

Initial Projects

Griffith University's vision for the implementation of a digital repository has been translated into ambitious goals for 2004. The focus has been directed towards a number of starter projects, covering a range of resource types and target audiences.

These projects have been selected on the basis of their ability to assist the Digital Repository Team to meet key objectives:

- Develop expertise in the Hive software
- Develop base metadata schema and then adapt as required for specialised content
- Develop database architecture to accommodate a range of differing content
- Develop appropriate workflows and business processes
- Identify policy, training and general implementation issues

In conjunction with the above objectives, the team is responsible for presenting options and identifying preferred solutions. The various starter projects contain quite disparate materials, each with its own set of challenges.

Previous Examination Papers

These were chosen because of their relative lack of complexity. The University already provides access to past exams (1998-2003) in PDF via its corporate web site. Quite coincidentally a staff member—as part of a different project—has created both MARC (Machine Readable) and XML (eXtensible Mark-up Language) records for each exam in the event that they are ever uploaded to the Library's OPAC (GriffLink). This set of data will be useful to initially test migrating already digitised objects and their accompanying data from several different systems to the Hive repository. This is envisaged to be a two-staged project, in which the team will later work with Exams and Timetabling to load current examination papers into the repository prior to invigilation.

Course Readings

As with past examinations, there already exists a corpus of digitised objects with descriptive records—in this case accessible via the GriffLink. However, unlike exams, digitised course readings inherently attract copyright issues. Under the Digital Agenda Act, there are strict guidelines as to how such material is to be communicated electronically. Therefore this project has provided the team with an opportunity to investigate Hive's Copyright Management Module.

It is envisaged that the module will enhance compliance by providing the following functionality:

- Flag items that are subject to copyright. When publishing an item, a user can specify that the item is not copyright, is subject to copyright, or contains or refers to a copyright work

- Record the item's copyright details, for example, its author and the type of copyright restrictions that apply to it
- Inform users that they are attempting to access copyright items
- Before allowing users to publish items, require them to acknowledge copyright restrictions and provide information about how they intend to use the items
- Record information about the status and usage of copyright items

For administrative reasons a "back room" area will be designed within the repository so that non-current digitised readings can be stored without invoking a violation of copyright.

Griffith University is currently trialling the implementation of a new Course Outline system so as to standardise on the provision of information and to create a more student-centred approach to that information. It is envisaged that in the near future course readings will be incorporated into that project, with a semi-automated updating function.

Art Objects

Griffith University's Queensland College of Art has an extensive collection (70,000) of art objects, particularly slides. A small percentage (2,000) of these exist in digital format and are accessible via a sophisticated, standalone Filemaker Pro database. However the system does not operate in a networked environment, thereby limiting potential cross-campus use. This project addresses such issues as storage (art images can have **very** large file sizes), graphical resolution, use of thumbnail surrogates, and once again copyright.

Whereas the majority of starter projects require very basic metadata, this project will test the requirement for specialised content. The team will investigate the use of the Visual Resources Association (VRA) Core Categories to support in-depth resource discovery by clients.

ePrints

The project involves the creation of a deposit collection of papers that showcase the research output of Griffith academic staff both before and after peer-reviewed publication. The objectives are similar to those that have been defined by the Group of Eight (2004) scheme:

- Create an archive of the publications produced at Griffith University
- Enhance access and contribution to internationally linked scholarship
- Increase awareness of and improve accessibility to Australian research, and
- Improve economics in scholarly communications

Teaching and Learning Objects

Within Griffith University the Educational and Products Services (EPS) unit within FLAS has highlighted the strategic importance of supporting the development of shared resources across disciplines. The challenge for EPS is that learning objects are developed by them and then may subsequently be deleted / removed from Learning@Griffith by new lecturers, thus precluding any re-use, recombining and/or sharing. Therefore there is a need to make the most of the considerable investment in a range of digital content to support teaching and learning.

Teaching and learning objects cover a wide range of formats, including Microsoft Word and Excel documents, PowerPoint slides, graphics and pictures, audio and video files, and HTML files. Many academics are keen to be able to search for existing learning objects for possible integration within their courses. Griffith's Digital Repository therefore is envisaged to enhance the storage and management of learning objects, including classification, re-use, presentation and improved searching functions. Ideally this should lead to academics combining objects to create their own courses designed to meet their learners' particular needs (CANDLE, 2004). Through collaboration the quality of courses and/or materials can be improved, along with learning outcomes.

Currently the Blackboard LMS is the principal repository for learning objects and that structure has led to some of the issues described above. In order to channel the upload function away from Blackboard and seamlessly into Hive, it is envisaged that simple interfaces will be built, using

'building blocks', ie applications that developers build to extend the Blackboard platform and to integrate Blackboard with external applications, content, or services.

Issues

There are a number of challenges which affect both the structure / architecture of a digital repository and its content.

Structure and Standards

The design of a digital repository needs to address the distinction between "archival resources", i.e. those that are likely to be long-term value, and "non-permanent" or transitory, i.e. those that have a very short lifecycle. Another challenge at Griffith is to design a repository structure that not only accommodates both structured and unstructured objects but also provides a mechanism for potential sharing of both types.

The ability to locate appropriate content is a critical success factor in creating reliable content and increasing its use. Metadata, ie a structured set of descriptive elements to describe an information resource, is the key in the identification, location and retrieval of those resources by end-users. The selection of a core metadata scheme—let alone specialised schema—has proved challenging because of the worldwide lack of consistency among digital repository initiatives. Griffith is investigating the IEEE's LOM (Institute of Electrical and Electronics Engineers' Learning Object Metadata) based on its adoption by the national ARROW Project (Harboe-Ree et al., 2003).

Culture, Politics and Practice

According to Sullivan (2004):

> Those involved in institutional repository projects have reported that the effort and organisational costs required to address repository policy, content management, and promotion to academic staff dwarf the technical implementation effort. The challenge . . . is not the technical implementation . . . but effecting the cultural change necessary for it to become an integral part of the activities of the institution.

This echoes the current perception at Griffith University. While the underlying objectives of digital repository projects are laudable, their actual implementation will need to address institutional core culture. Change management is therefore the single largest challenge facing the Digital Repository Team.

Academic staff are understandably concerned about intellectual property and the perceived threat by exposure of their resources to a wider audience. Another challenge is to demonstrate how existing content can be re-used to support learning (Corti, 2003). At what level of granularity do learning objects need to be stored to be capable of being re-used, combined and/or shared? And as Campbell, Blinco & Mason (2004) point out, along with defining incentives for academics to use repositories, institutions such as Griffith University need to define roles and responsibility. Who will be responsible for metadata creation? Can Griffith realistically expect academic staff to perform this task?

Even in the preliminary stages of the Griffith Digital Repository Project, it is already clear that formal University-wide policies will need to be written to support the above projects. Issues such as intellectual property, archival retention period (s), granularity and copyright management need to be addressed.

The Future

Once prototypes have been developed for each of the starter projects, these will become pilot projects with academic staff involving faculty librarians, educational designers, and multimedia developers to test solutions and obtain feedback. The Digital Repository Team is keen to work with a School or Centre in managing their collection(s) of digital resources before rolling out the repository across all educational units.

Given the importance of resource discovery, it is anticipated that a new Federated Search service will be layered on top of the Hive repository to further enhance access. The Team will continue to maintain close contact with other institutional and national digital repository projects in order to share ideas and experiences. Griffith University is aware of the benefits of ensuring there is interoperability across the higher-education sector.

Griffith University is striving to create a resource-rich, integrated learning environment that is fully exploited by both students and faculty. A digital repository provides a flexible and discipline-independent mechanism for storing and managing digital objects, thus enhancing integrating learning environments.

References

Campbell, L., Blinco, K. & Mason, J. (2004). *Repository management and implementation*. [Online]. Available: http://members.imsglobal.org/forum/ims/dispatch.cgi/f.altilabtech/showFile/100061/d20040712142835/No/Altilab04-repositories.pdf [10th August 2004].

CANDLE (2004). *The CANDLE project – Collaborative and Network Distributed Learning Environment*. [Online]. Available: http://www.candle.eu.org/ [12th August 2004]

Corti, L. (2003). *Re-purposing survey data sources for teaching and learning*. [Online]. Available: http://www.iassistdata.org/conferences/2003/presentations/E1_Corti.ppt [10th August 2004].

Group of Eight. (2004). *Statement on open access to scholarly information. [Online]*. *Available*: http://www.go8.edu.au/news/2004/Go8%20Statement%20on%20open%20access%20to%20scholarly%20information%20May.pdf [15 September 2004].

Harboe-Ree, C., Treloar, A., & Sabto, M. (2003). *ARROW: Australian Research Repositories Online to the World*. [Online]. Available: http://eprint.monash.edu.au/archive/00000046/ [16th August 2004]

Leslie, S. (2004*). 'Institutional Digital Repositories' and 'Learning Object Repositories' - What's the Difference?* [Online]. Available: http://www.edtechpost.ca/mt/archive/000485.html [16th August 2004]

Long, P. (2004). *Learning Object Repositories, Digital Repositories, and the Reusable Life of Course Content* [Online]. Available: http://www.syllabus.com/article.asp?id=9258 [16th August 2004]

McLean, N. & Lynch, C. (2004). *Interoperability between Library Information Services and Learning Environments – Bridging the Gaps*. [Online]. Available: http://www.imsglobal.org/digitalrepositories/CNIandIMS_2004.pdf [16th August 2004]

Peters, T. (2002). Digital repositories: individual, discipline-based, institutional, consortial or national? *Journal of Academic Librarianship*, 28(6), 414-417.

Sullivan, S. (2004). *New models of research publishing*. [Online]. Available: http://ausweb.scu.edu.au/aw04/papers/refereed/sullivan/paper.html [10th August 2004].